Technical Disclosure Commons

Defensive Publications Series

July 03, 2017

Delivery Of Content Based on Search Entity-Action Pairs In a Voice Activated Computing System

John D. Lanza
Foley & Lardner LLP

DELIVERY OF CONTENT BASED ON SEARCH ENTITY-ACTION PAIRS IN A VOICE ACTIVATED COMPUTING SYSTEM

Voice activated computing systems provide a user with content or services in response to voice commands spoken by the user. Such systems can capture voice commands from a user, process the voice commands to determine requests and keywords in the voice commands, and provide the user with content or services related to the determined requests and keywords.

As discussed herein, a voice activated computing system processes the voice commands to identify search requests and trigger keywords in the voice commands. Based on the search requests and the trigger keywords, the system can identify search entities, and associated actions. For example, if the voice command provided by the user were: "I would like to go to a restaurant today," the system can identify "restaurant" as the search entity, and one or more of generating a list of names of nearby restaurants, making reservations at a restaurant, providing reviews of the restaurant, estimating travel time to a restaurant, and the like. The search entity and the actions can be determined based on techniques such as heuristics and knowledge graphs. The system can then form one or more search entity-action pairs, such as for example, a "restaurant" and "making reservations." Third-party content providers, such as restaurants, and reservation service providers, can compete to provide their content to the user based not on the search request "restaurant" alone, but on the search entity-action pair. That is, the third-party content providers can enter an auction and bid for providing their content over others. The selected third-party content can then be provided to the user. In some instances, each content available for a search entity-action pair can have an associated quality score, which can be representative of a likelihood of the user accepting the content if presented to the user. Based on a combination of the auction bid and the quality score, the system can select a content item and generate an

audio file that provides the content item to the user. For example, the system may send a voice response to the user stating: “Restaurant XYZ has reservations available in 1 hour, would you like to go ahead and make the reservation?”

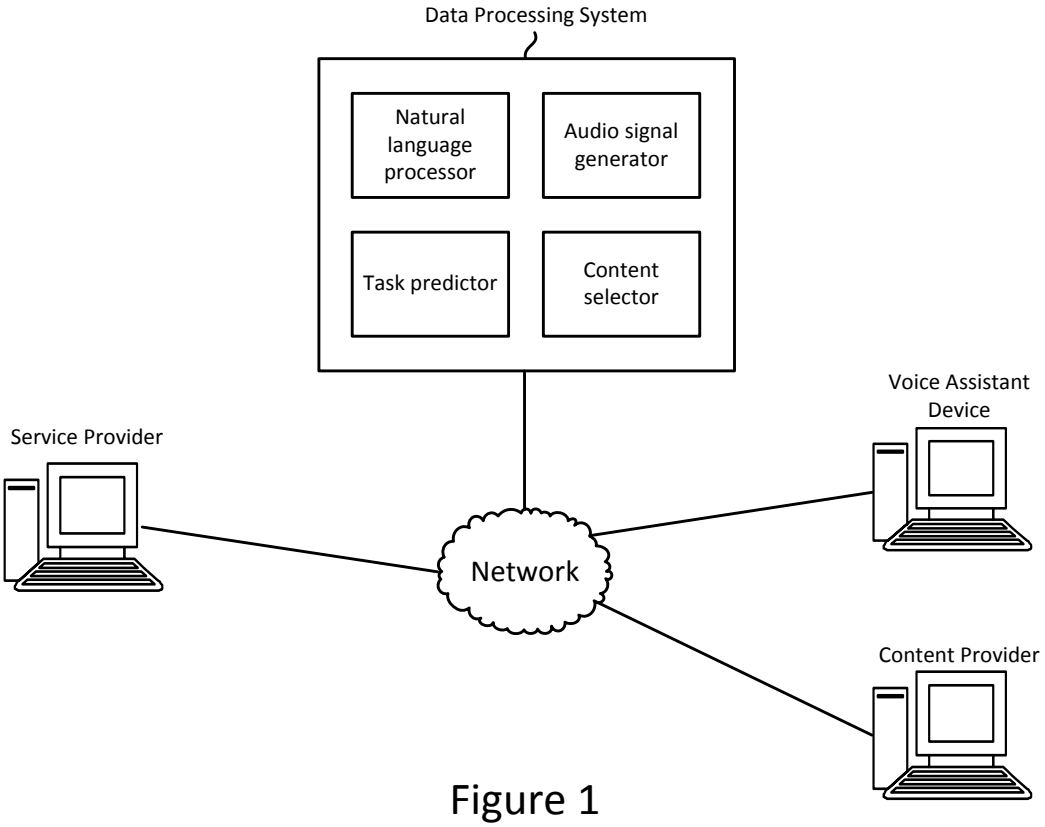

Figure 1

Figure 1 shows an example voice activated computing system. The system includes a voice assistant device, a service provider, a data processing system and a content provider communicating over a network. The voice assistant device can be a device accepts voice commands, and provides audio or visual output. The voice assistant can include one or more mics and cameras, such that voice commands received by the user are converted into corresponding audio signals. The voice assistant can send the audio signals to the data processing system and the service provider. The voice assistant device also can receive data such as audio signals or video signals from the data processing system or the service provider.

The voice assistant device also can include audio speakers that can convert the audio signals received from the data processing system or the service provider into sound.

The data processing system can process voice commands received from the voice assistant device. The data processing system includes a natural language processor, an audio signal generator, a task predictor, and a content selector. The natural language processor is capable of processing voice commands included in the audio signals received from the voice assistant device. The natural language processor can convert the audio signals into recognized text by comparing the audio signals against a stored, representative set of audio waveforms, and choosing the closest matches. The representative waveforms are generated across a large set of users, and can be augmented with speech samples. After the audio signals are converted into recognized text, the natural language processor can match the text to words that are associated, for example via training across users or through manual specification, with actions that the data processing system can serve. Basically, the natural language processor identifies requests and trigger words in the converted text, based on which the natural language processor can determine the content and actions to be carried out.

The task predictor can predict tasks or actions based on the converted text, and in particular by identifying requests and trigger keywords in the converted text. The task predictor also can predict sponsored content related to the converted text.

The content selector can select content, such as services or actions to be offered to the user based on the actions identified by the task predictor. The content selector also can communicate with the content server to obtain sponsored content identified by the task predictor. The audio signal generator can generate audio signals based on the actions selected by the

content selector. The audio signals can be representative of voice responses or voice instructions provided to the user in response to the voice commands.

The service provider can provide one or more service to the user. For example, the service provider can provide weather forecast, traffic conditions, and the like. The service provider can communicate with the data processing system to provide information related to the requested service. Alternatively, the service provider can communicate directly with the voice assistant device independently of the data processing system. The service provider can also include a natural language processor, similar to the one discussed above in relation to the data processing system, to convert user voice commands into text, and identify requests and keywords to determine the services requested by the user.

Referring again to the voice command example mentioned above, the user can speak the voice command "I would like to go to a restaurant today," to the voice assistant device. The mics at the voice assistant device can convert the voice commands into corresponding audio signals, which are be transmitted by the voice assistant device to the data processing system over the network.

At the data processing system, the natural language processor processes the audio signal received form the voice assistant device and identify a request for a "restaurant." The natural language processor also can identify a trigger keyword "go" or "to go to," which can indicate interest in going to a restaurant, travel to the restaurant, and the like. Based on the request and the trigger keywords, the natural language processor can identify one or more search entities. A search entity can include keywords for which merchants can provide sponsored content. For example, merchants can bid for providing their sponsored content in response to a keyword. An auction can be used to determine which of the merchants sponsored content is to be presented to

the user. In the above example, the natural language processor can identify "restaurant" or "restaurant in zip code 12345," as search entities.

In some instances, the natural language processor, or the system, can utilize a knowledge graph to determine the most appropriate or a list of most appropriate search entity. A knowledge graph generally refers to a structured graph used in semantic searching where each node in the graph represents a unique search entity and links between entities represent the relationships between them. Each node may also include any number of facts for a search entity. For example, a node for the "Sheriff Bob" movie may include facts about the movie, such as its release date, list of characters, list of actors, or the like. The system may search the knowledge graph by comparing facts and relationships for different search entity nodes to the request and trigger keywords to identify a matching search entity. The system may also use a machine learning model to identify a search entity. Such a model may, for example, analyze the search entities matched using a heuristic search of a knowledge graph, match corrections supplied by third-party content providers, analyze vertical information associated with campaigns, or the like, to identify the most likely search entity corresponding to the request and trigger keywords.

The natural language processor also can identify action associated with each search entity. An action can refer to actions performed regarding the search entity. For example, actions for the search entity "restaurant" can include generating a list of names of nearby restaurants, making reservations at a restaurant, providing reviews of the restaurant, estimating travel time to a restaurant, and the like.

The content selector can be configured to select third-party content, provided by a service provider, based on an entity-action pair. For example, a third-party content provider can specify that they wish to provide sponsored content whenever the entity-action pair including

"restaurant" and "reservations" identified in a voice command. A search service may also conduct a content auction in which different third-party content providers compete for the ability to include their respective third-party content when a particular entity-action pair is identified. For example, multiple restaurant reservation providers may compete to provide restaurant reservations in response to the entity-action pair of "restaurant" and "reservations." In another example, the action can include "ride" associated with the entity "restaurant." The content selector can select a third-party content provider, such as ride sharing service providers, taxi service providers, and the like, from among a number of third-party content providers competing to provide their products and services whenever the entity-action pair of "restaurant" and "ride" is identified.

In some instances, the system can identify additional search entities that are related to the first search entity identified in the voice command received by the user. For example, the search entity for information about restaurants can be related to a search entity for information about Italian restaurants. Determining if the search entities are related can include performing a lookup in a log storing the occurrence of each search entity. The system can identify the number of times the first search entity was requested and the number of times the second search entity was requested. The system can determine the number of times the request for the first search entity and the second search entity originated from the same source. If the number of times the request for the first search entity and the second search entity originated from the same source occur more than a predetermined number of times, the system can determine that the two search entities are related. The request for the first search entity and the request for the second search entity can be counted as originating from the same source if the requests originate within a single session or within a predetermined amount of time of one another. The first search entity can be

determined to be related to the second search entity if the number of times the search for the second search entity and the search for the first search entity originated from the same source is above the predetermined threshold.

The system may also determine a quality score for the content items identified for the search entity-action pair. A quality score can represent a likelihood that a user would proceed with the content item, if the content item were to be presented to the user. The quality score can be used in addition to, or in lieu of, the third-party content provider's auction bid to determine whether or not the providers' content item will be presented to the user. A third-party content provider that has the highest auction bid may nonetheless be passed over for a content selection if the provider has a low quality score. In some instances, entity-action pairs may also have associated reserve price set, such that one or more bids must be above the reserve price before its associated content will be selected for presentation to the user.

Once the content selector selects a content item, the selected content item can be presented to the user. For example, the selected content item may include the name of a restaurant "Restaurant XYZ" and information that reservations are available in an hour. The audio signal generator can generate an audio signal to present this information to the user. For example, the audio signal generator can generate an audio signal with a voice message: "Restaurant XYZ has reservations available in 1 hour, would you like to go ahead and make the reservation?" Of course, other voice messages providing the same information to the user can be generated. The user, in response to the voice message, may provide additional voice commands indicating that it would like to make the reservations, or that it would like to hear additional options. If the user would like to hear additional options, the content selector may provide reservation information from other content item providers that were not previously selected.

Abstract

This document describes a technique for processing voice commands received by a voice activated computing system to identify search requests and trigger keywords in the voice commands. Based on the search requests and the trigger keywords, the system can identify search entities, and associated actions. The search entity and the actions can be determined, for example, based on techniques such as heuristics and knowledge graphs. The system can then form one or more search entity-action pairs. Third-party content providers can compete to provide their content to the user based not on the search request alone, but on the search entity-action pair. That is, the third-party content providers can enter an auction and bid for providing their content over others. The selected third-party content can then be provided to the user. In some instances, each content available for a search entity-action pair can have an associated quality score, which can be representative of a likelihood of the user accepting the content if presented to the user. Based on a combination of the auction bid and the quality score, the system can select a content item and generate an audio file that provides the content item to the user.